

本周周报（2013.10.28-2013.11.23）

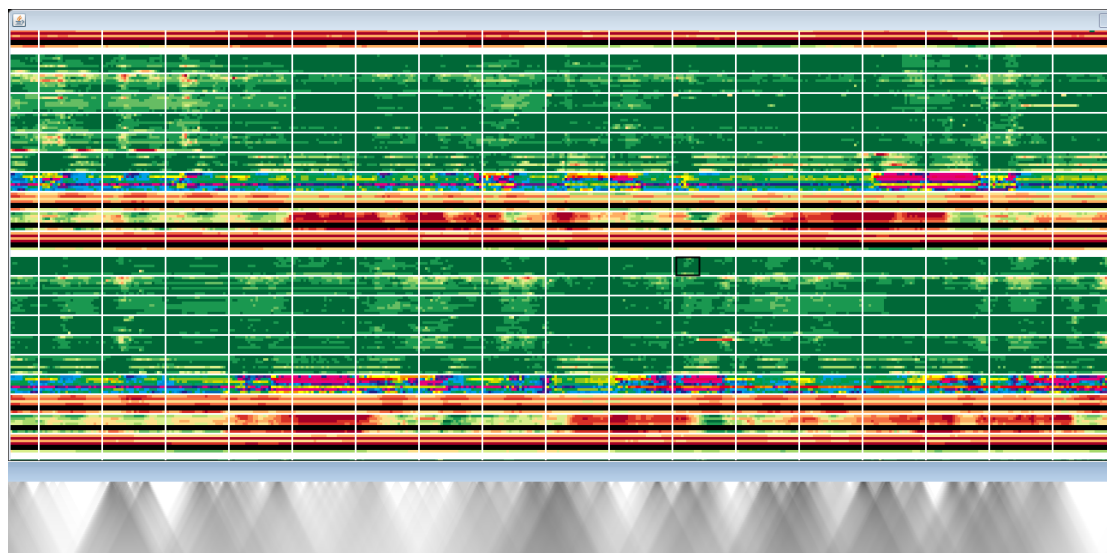
郭方舟

本周工作

1. 空气污染数据可视化

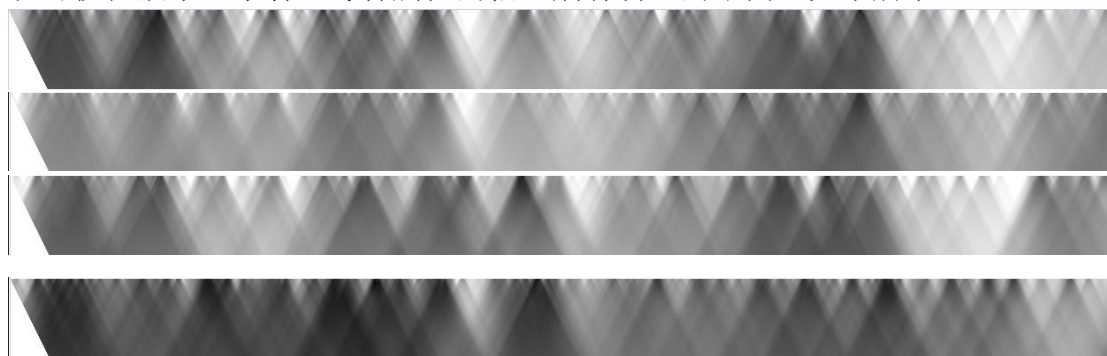
由于本周斐然师兄出去开会，所以没有确定下来自适应表达的实现方式。与师兄讨论过后，认为目前使用信息熵的方法可以找出一些事件，但是信息熵是一种基于统计的方法，会丢失掉时序上的信息，因此考虑将信息熵当成一个特征值，再尝试用其他方法来对数据处理，算出其他的特征值来综合确定事件，比如 sift、ck 算法等等。

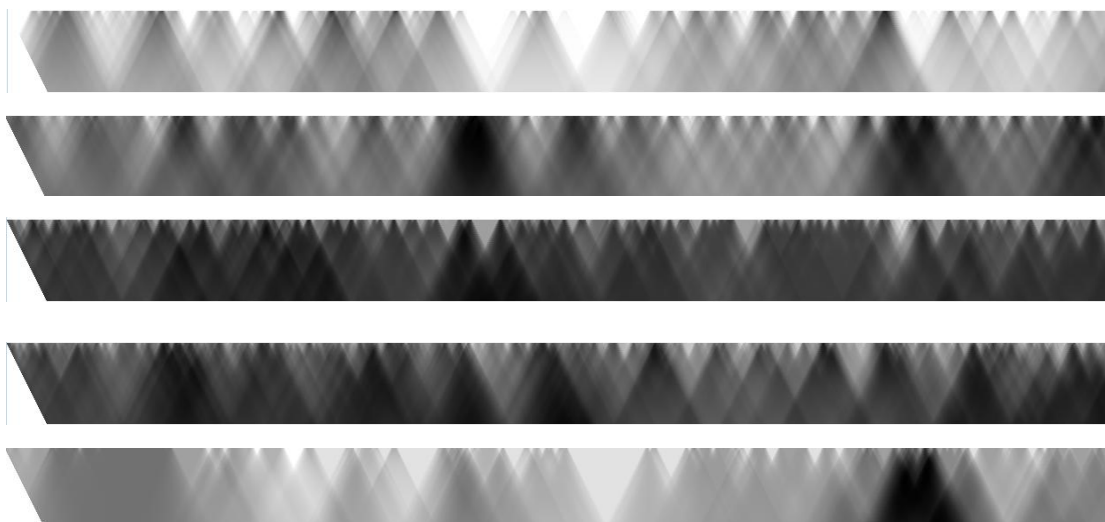
同时，在计算相关性时，如果对每个属性，每个传感器的数据加窗并计算数据之间的相关性的话，计算量非常大而且会产生巨量的数据。所以在识别出事件之后再计算事件与事件间的相关性可能更好。



本周计算了不同属性的信息熵并进行了可视化，同时也计算了不同属性，不同传感器的信息熵，但是还没有进行可视化。

从可视化效果上来看，每种属性的信息熵有明显的不同，如下所示：

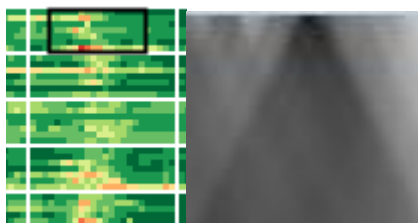




从上面的图中可以看到，前五条比较相似，而后四条不是很相似。前五个属性分别是 CO，NO₂，PM₁₀，SO₂，NO_x，全部都是代表空气污染情况的属性，所以他们的信息熵的模式比较近似，峰值出现的区域也比较接近；而后四个属性分别是 WS，TEMP，HUMD，PRESS。后四个属性的信息熵图之间的相似度不高，与前五个属性的相似度也不高，这可能隐含了一些可以挖掘的事情，但是还没有仔细讨论。

从目前的结果来看，使用信息熵是可以找到一些事件的，列举如下：

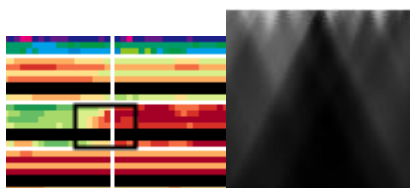
1) 短时间事件



2) 波段事件



3) 突变事件



4) 空事件



这四种事件是目前从熵图中找到的几个代表性的事件。

2. 完成马克思与当代的课程作业, 包括一篇读书报告和一篇课程论文

下周工作

1. 空气污染数据的可视化

阅读关于 `sift`、`ck` 等算法的论文, 找出使用的算法并实现。

与斐然师兄讨论自适应表达的问题, 并开始实现。